



UNIVERSIDAD
COMPLUTENSE
MADRID

Proyecto de Innovación y Mejora de la Calidad Docente
Convocatoria 2019/20

Nº 189

*Estudiantes y Campus Virtual. Utilidad del Learning Analytics para identificar
luces y sombras y caminar hacia metodologías adaptativas.*

M^a Pilar Álvarez Vázquez
Facultad de Medicina
Departamento de Biología Celular

1. OBJETIVOS PROPUESTOS EN LA PRESENTACIÓN DEL PROYECTO

El proyecto mantiene como objetivo principal:

- Proporcionar información a los docentes para la implementación de su metodología docente digital de acuerdo al análisis de los registros del campus virtual (CV) de la UCM y posterior interpretación sobre el comportamiento de sus estudiantes.

Este objetivo se estructura en los siguientes objetivos específicos:

1. Realizar un análisis descriptivo de los usos que hacen los estudiantes de cada una de las asignaturas virtualizadas en tres grados de CC de la Salud (Medicina, Fisioterapia y Podología), y una de del Máster Universitario en Investigación en Cuidados de la Salud (MUICS)
2. Realizar un análisis comparativo horizontal y vertical del uso que hacen los alumnos del CV:
 - Horizontal: comparar comportamientos en asignaturas del mismo curso (2º en tres titulaciones distintas de grado.
 - Vertical:
 - comparar comportamientos en asignaturas de distinto curso (2º y 4º de la misma titulación (Grado en Podología),
 - comparar comportamientos entre asignaturas del grado y una asignatura del Máster Universitario en Investigación en Cuidados de la Salud (nivel educativo superior), todas de la misma facultad.
3. Establecer, si fuera posible, distintas tipologías de estudiantes de acuerdo a su comportamiento y uso de las asignaturas virtualizadas analizadas.
4. Correlacionar los posibles clústeres de estudiantes encontrados en dicha asignatura virtualizada con otros parámetros, como su rendimiento académico.
5. Realizar el análisis inverso: partiendo del rendimiento académico, establecer distintas categorías y analizar si cada una de ellas se comporta en el CV de manera similar o distinta respecto de las otras. Establecer las posibles diferencias respecto del análisis anterior.
6. Proporcionar al docente información sobre la usabilidad, dinamismo y nivel de actividad desempeñados por los estudiantes en su asignatura virtualizada, mediante el establecimiento de los indicadores que nos permitan poner en marcha propuestas educativas innovadoras para uso de la plataforma Moodle.

2. OBJETIVOS ALCANZADOS

Se han trabajado todos los espacios propuestos correspondientes al primer cuatrimestre o que hubieran terminado la evaluación en primera convocatoria dentro del plazo de presentación de la memoria. En la Tabla 1 se indican los espacios del CV incluidos en el PIMCD. Los dos últimos, en tono sepia, están pendientes de análisis ya que debido a los cambios en el calendario académico a consecuencia del COVID19, las nuevas fechas de evaluación no hacen posible que se incluyan en la memoria.

Tabla 1. Espacios virtuales incluidos en el PIMCD 2019. En sepia los que no se han podido analizar debido al retraso en el calendario académico

Nombre del espacio	identificador	Curso	Titulación: Máster en
EVIDENCIAS MUICS	18-107910	2018/19	Investigación en Cuidados de la Salud
EVIDENCIAS MUICS	19-107910	2019/20	Investigación en Cuidados de la Salud

Nombre del espacio	identificador	Curso	Titulación: Grado en
Anatomía Humana III	19-141227	2019/20	Fisioterapia
Biomecánica	19-171904	2019/20	Podología
Salud Laboral	19-473738	2019/20	Podología
PRÁCTICAS Organografía Grupo IA 19/20	seminario-invest-5914-21	2019/20	Medicina
Ortopodología I 2019-20	19-171914	2019/20	Podología
ORGANOGRÁFIA MICROSCÓPICA HUMANA (GRUPO IA)	19-93062	2019/20	Medicina

El trabajo en equipo interdisciplinar ha permitido progresar en el desarrollo del proyecto, y profundizar en los resultados del PIMCDpimcd2018-20 para su difusión y publicación.

3. METODOLOGÍA EMPLEADA EN EL PROYECTO

- Encuesta a los estudiantes sobre CV, acceso a herramientas web 2.0 y bibliotecas, conocimiento de TIC y características socio-demográficas. Este curso se ha utilizado Google Forms para que rellenaran la encuesta *online* a fin de simplificar el trabajo de la transcripción de datos por parte del PDI y PAS.
- Fuentes de datos de los estudiantes analizados: encuesta, *logs* descargados de *Moodle* y rendimiento académico.
- Extracción de los *logs* de cada espacio virtual a analizar, descarga de calificaciones y obtención de las rutas de cada recurso y actividad. Las variables consideradas fueron número de: accesos a CV, accesos a recursos y a URL externas, autoevaluaciones y tareas realizadas, consultas a foros y foros.
- Estandarización de archivos, depuración y tratamiento anonimizado de los *logs*.
- Análisis con R. Creación de un aplicativo que permite cruzar resultados y gráficos con un número ilimitado de asignaturas y hasta 30 calificaciones en cada una, sistematizando las opciones de tratamiento de datos.
- Elaboración de una Declaración de Tipo de Datos (DTD) que permite documentar la estructura del documento XML.
- Análisis estadístico encuesta. Variables cualitativas mediante frecuencias, porcentajes y tablas de contingencia para visualizar la relación entre pares de variables, posteriormente analizada por estadístico exacto de Fisher.
- Análisis estadístico de uso de CV comparado con rendimiento académico. Estudio de cada asignatura. Correlaciones no paramétricas de Spearman entre calificaciones y accesos a CV, recursos y actividades. Se obtuvieron árboles de decisión con dos criterios de corte (minimización de la significatividad mediante el estadístico de Fisher-Snedecor y minimización de la varianza) para relacionar rendimiento académico con segmentos de alumnos con una actividad similar en CV.
- Análisis de correspondencias múltiples con las variables cualitativas seguido de un análisis de conglomerados jerárquico aplicado a las puntuaciones factoriales y utilizando la distancia de Ward. Se procede a un corte en cuatro conglomerados que representan el 48% de la variabilidad original. Se caracterizan los conglomerados mediante el estadístico Chi-2 destacando aquellas características con una mayor o menor representatividad que en el conjunto total. Se proyectan en los dos primeros ejes factoriales tanto las variables como los individuos, añadiéndose en los mismos como variables suplementarias las calificaciones.
- Visualización de datos: Curvas de frecuencias de acceso a CV, recursos y actividades en el curso académico; Diagramas de barras según acceso a CV, recursos y actividades día/semana; Mapas de calor según acceso a CV, recursos y actividades según día y hora;

Diagramas de barras mostrando la distribución de las variables para visualizar diferencias entre grupos; Grafos para representar árboles de decisión; y Gráficos de estrella multivariantes de Kiviat para representar accesos promedio a CV, recursos y actividades según segmentos de calificaciones.

4. RECURSOS HUMANOS

En este PIMCD han participado PDI, PAS y una estudiante.

- A. Álvarez Méndez, M.P. Álvarez Vázquez y J. Cristóbal Barrios han realizado búsquedas bibliográficas y discutido los trabajos de otros colegas.
- PDI ha elaborado la encuesta de satisfacción de alumnos sobre el CV y Y. Fuentes creó un formulario en Google Forms. El procesamiento de los datos corrió cargo de C. Bravo, del Servicio de Apoyo a la Docencia de la UCM.
- PDI, cada profesor ha proporcionado sus listas de clase, las fechas importantes en la materia, así como el rendimiento académico de sus estudiantes.
- J. Cristóbal Barrios ha depurado y analizado de forma anonimizada todos los *logs* de *Moodle* de las asignaturas y espacios analizados, los listados de estudiantes y de calificaciones, obteniendo distintos tipos de gráficas. Ha elaborado un nuevo manual técnico.
- C. Bravo, del Servicio de Apoyo a la Docencia de la UCM ha realizado análisis estadísticos de asignaturas y análisis multivariante de 8 espacios del PIMCD2018-20. En la preparación de los datos y discusión de los mismos participaron A. Álvarez, M.P. Álvarez y M.T. Angulo.
- Además, analizó estadísticamente los datos agregados de cada asignatura, realizado distintos tipos de gráficas, correlaciones y árboles de decisión.
- El conjunto del PDI y PAS se ha encargado de debatir las características de la materia, de planificar las etapas de la investigación, de discutir la categorización de los registros, de analizar resultados y extraer conclusiones, y de elaborar distintos trabajos y comunicaciones a congresos.
- PDI y PAS han contribuido a la redacción de la presente memoria.

5. DESARROLLO DE LAS ACTIVIDADES

En la primera parte del curso 2019/20, el equipo se centró en:

1. Análisis “inverso” con R de los espacios del PIMCD2018-20, esto es, partiendo de grupos de alumnos según su calificación, establecer si dichos grupos presentaban o no comportamientos diferenciales.

En esta fase inicial se realizaron nuevas gráficas de estrella. Podemos apreciar diferencias interesantes que nos llevan a concluir que esta aproximación nos parece mejor que la realizada en el PIMCD2018-20. Presentamos a continuación una gráfica comparativa de ambos sistemas (Fig. 1) y otra que muestra diferentes opciones según la calificación que interese analizar (Fig. 2).

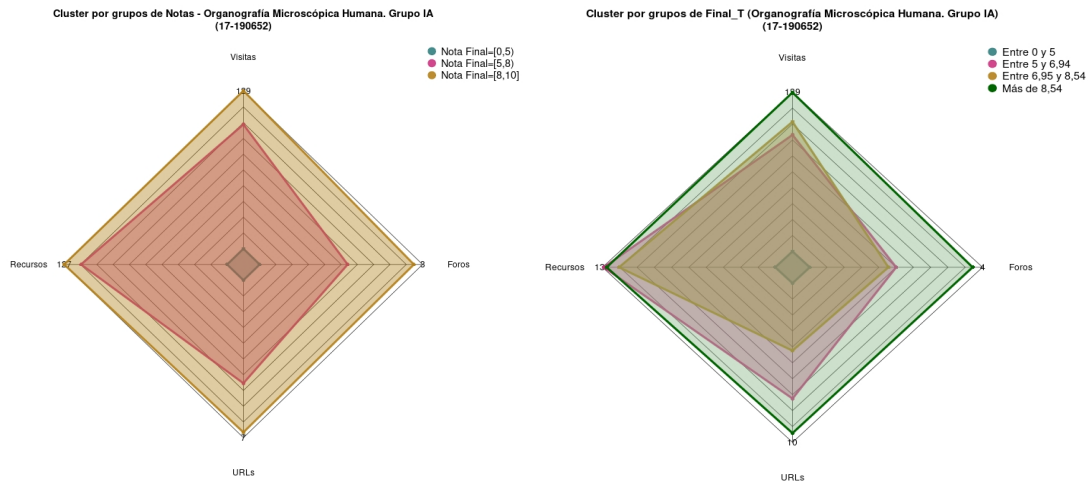


Figura 1. Gráficos de estrella multivariable de Kiviat correspondientes al espacio de prácticas de la asignatura OMH, curso 2017/18. A la derecha se representa la actividad registrada por los 4 clústeres de estudiantes hechos según su nota final.

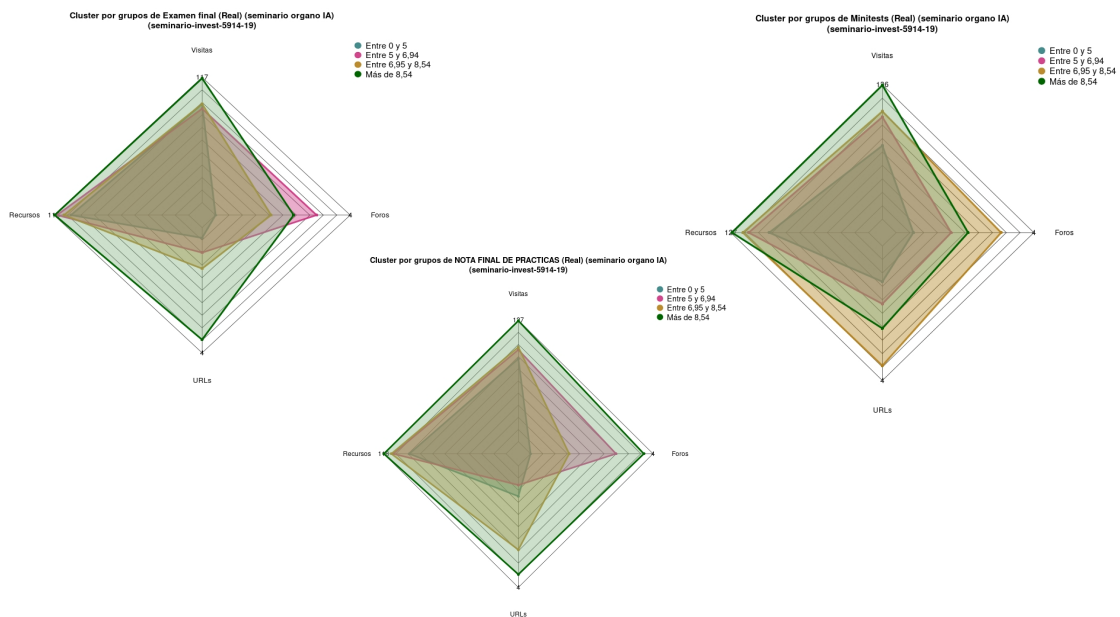


Figura 2. Gráficos de estrella multivariable de Kiviat correspondientes a la actividad registrada en el espacio de prácticas de la asignatura OMH, curso 2018/19, de cuatro clústeres de estudiantes según tres tipos de notas: minilests, examen final y nota final de prácticas.

2. Elaboración y discusión de resultados del PIMCD2018-20 de cara a su publicación. Estas tareas fueron llevadas a cabo por PDI, PAS y la estudiante N. Molina.
Los resultados principales obtenidos en el PIMCD2018-20 han sido objeto de varias comunicaciones presentadas a congresos y una publicación.
 - a) Álvarez Vázquez M.P., Álvarez Méndez A., Angulo Carrere M.T., Molina Hernández N., Bravo Llatas M.C., Cristóbal Barrios J. *Learning Analytics: Herramienta para conocer mejor a nuestros estudiantes*, presentada en el XXIV Congreso de la SEDEM, Salamanca, 6-8 noviembre 2019.

- b) Angulo Carrere M.T., Álvarez Méndez A., Molina Hernández N., Bravo Llatas M.C., Cristóbal Barrios J., Álvarez Vázquez M.P. *Uso del campus virtual y rendimiento académico*, presentada en XXIV Congreso de la SEDEM, Salamanca, 6-8 noviembre 2019.
- c) M.P. Álvarez Vázquez, A. Álvarez Méndez, M.T. Angulo Carrere, J. Cristóbal Barrios, C. Bravo Llatas. *Learning Analytics in Human Histology reveals different student clusters and different academic performance*, presentada en INTED 2020 y publicada en Proceedings of INTED2020 Conference 2nd-4th March 2020, Valencia, Spain 2020, pp. 66-72. ISBN: 978-84-09-17939-8.
- d) A. Álvarez-Méndez, T. Angulo Carrere, J. Cristóbal Barrios, C.Bravo-Llatas, M.P. Álvarez Vázquez. *Application of Data Mining in Moodle platform for the analysis of the academic performance of a compulsory subject in university students*, presentada en INTED 2020 y publicada en Proceedings of INTED2020 Conference 2nd-4th March 2020, Valencia, Spain, 2020, pp. 984-992. ISBN: 978-84-09-17939-8.
- e) M.P. Álvarez Vázquez, A. Álvarez-Méndez, C. Bravo-Llatas, J. Cristóbal Barrios, M.T. Angulo Carrere, *Tipologías de estudiantes de Fisioterapia según el uso que hacen del campus virtual*, en Revista de Innovación Docente Universitaria 12 (2020), pp. 74-81. DOI: <http://dx.doi.org/10.1344/RIDU2020.12.8>.

Todos los trabajos anteriormente citados, se han subido a Eprints UCM. Por último, dos trabajos más están siendo evaluados para publicación.

3. Análisis con R de los registros de la asignatura Evidencia (E) del curso 2018/19 de la profesora Y. Fuentes, según procedimiento descrito por J. Cristóbal.
4. Análisis comparado del comportamiento en CV de los alumnos de OMH del curso 2018/19, realizado por C. Bravo. La razón de incluir aquí este estudio es que la 2ª convocatoria se realizó en julio, fuera del plazo de PIMCD2018-20, pero consideramos importante saber si existen diferencias entre los alumnos que aprobaban en la 1ª o en la 2ª convocatoria. Los resultados revelan importantes diferencias entre los alumnos que aprobaron la parte teórica por parciales y los que fueron al examen final, con diferencias significativas entre ambos grupos en cuanto a visitas, acceso a recursos y a foros (Tabla 2).

Tabla 2. Contraste de distribuciones entre las calificaciones teóricas de OMH en las dos convocatorias

Resumen de prueba de hipótesis			
	Hipótesis nula	Prueba	Sig.
1	La distribución de Visitas es la misma entre las categorías de Convocatoria.	Prueba U de Mann-Whitney para muestras independientes	,000
2	La distribución de Recursos es la misma entre las categorías de Convocatoria.	Prueba U de Mann-Whitney para muestras independientes	,000
3	La distribución de Tareas es la misma entre las categorías de Convocatoria.	Prueba U de Mann-Whitney para muestras independientes	,001
4	La distribución de Foros es la misma entre las categorías de Convocatoria.	Prueba U de Mann-Whitney para muestras independientes	,043
5	La distribución de URLs es la misma entre las categorías de Convocatoria.	Prueba U de Mann-Whitney para muestras independientes	,379

Se muestran significaciones asintóticas. El nivel de significación es de ,0

En el segundo cuatrimestre, el trabajo del equipo giró en torno a:

1. Análisis estadístico “inverso” de espacios analizados en el PIMCD2018-20, por C. Bravo. Se procedió a analizar los correspondientes a OMH, teoría y prácticas. Los resultados del análisis no paramétrico indican en la teoría usos significativamente diferentes entre los alumnos de la categoría *Suspense* y los de las demás categorías en la convocatoria de junio, mientras que en la parte práctica se encontraron diferencias estadísticamente significativas en la categoría *Sobresaliente* respecto del resto de categorías solo en cuanto a la nota de los minitests.
2. Una vez concluido el primer cuatrimestre así como la parte práctica de OMH, se procedió a la descarga de los registros de cada uno de los espacios para que el técnico J. Cristóbal procediera a su depuración y análisis. En los 6 espacios analizados se procesaron un total de 127.764 *logs*. Asimismo se facilitaron los ficheros correspondientes a las fechas de inicio y fin, calificaciones, identificadores, fechas clave de cada una y actividades llevadas a cabo en el CV.

Los principales hallazgos obtenidos se resumen a continuación.

En todos los espacios se repite la estrecha relación entre la actividad en el CV de los estudiantes y los calendarios académicos, fechas de entregas, autoevaluaciones y exámenes.

En la asignatura de Biomecánica (BM) se procesaron un total de 23.654 *logs*. Los resultados han sido claramente mejores que en el curso 2017/18, analizado en el PIMCD2018-20. Por una parte, el rendimiento académico ha mejorado en la convocatoria de febrero pasando los aprobados del 20 al 45% y disminuyendo el número de no presentados del 20 al 10%. Por otra parte, el comportamiento de los estudiantes en el CV se muestra menos errático, con un aumento notable en el acceso a recursos y a URLs. Ambos cambios están relacionados como se puede constatar al comparar la actividad media en los clústeres según sus calificaciones en primera convocatoria, examen de teoría y examen de prácticas. En el caso del examen teórico se observa que los estudiantes con las notas más altas se relacionan con un mayor uso del CV (visitas), utilización de recursos y la realización de las autoevaluaciones, las cuales se relacionan directamente con los contenidos teóricos (Fig. 3). En el caso del examen práctico las calificaciones mejoran y se relacionan con el acceso a URLs, recursos especialmente asociados a los contenidos de prácticas.

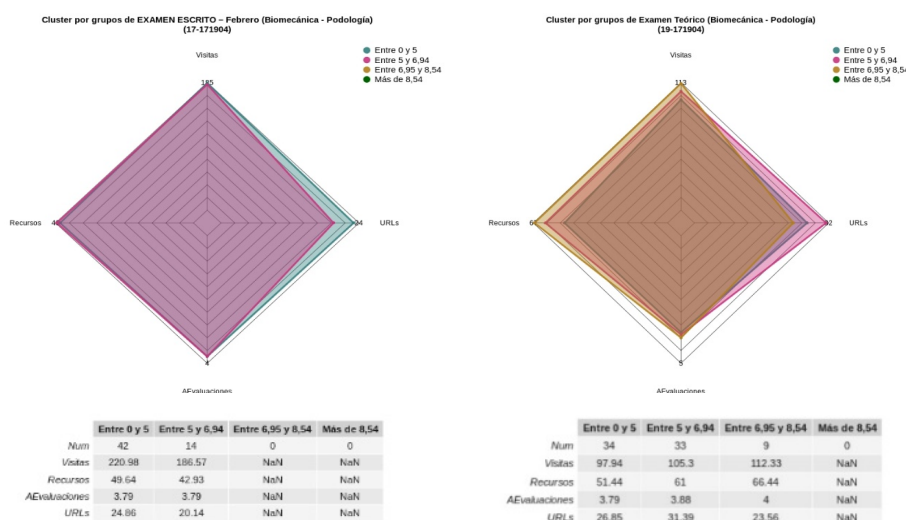


Figura 3. Comparativa de la actividad media en el CV de los clústeres de estudiantes de BM según su nota en el examen teórico en los cursos 2017/18 (izqda.) y 2019/20 (dcha.)

Frente a esta mejoría, los resultados fueron peores en Anatomía Humana III (AH) en comparación con los obtenidos en el PIMCD2018-20 tanto en lo que se refiere a los registros de actividad como al rendimiento académico. Se puede observar en la Figura 4 que la escala en el mapa de calor disminuye a la mitad (600 vs 300). El número de *logs* procesados en AH fue de 21.246.

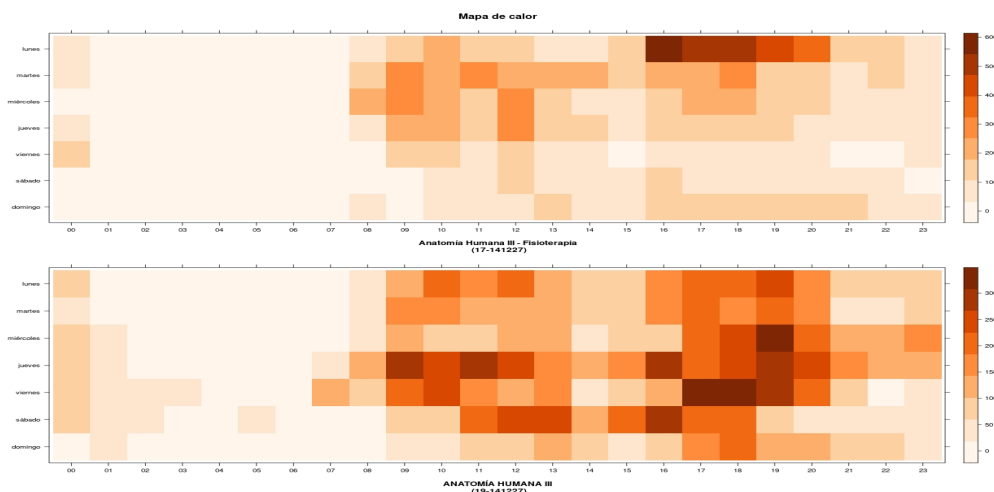


Figura 4. Comparativa de los mapas de calor en la asignatura AH (arriba curso 2017/18, abajo curso 2019/20)

En la asignatura de Salud Laboral (SL) llama la atención el fuerte descenso en el uso del CV en comparación con el curso 2018/19 (Fig. 5) y que se manifiesta también en el total de *logs* procesados: 8.915 vs 19.957.

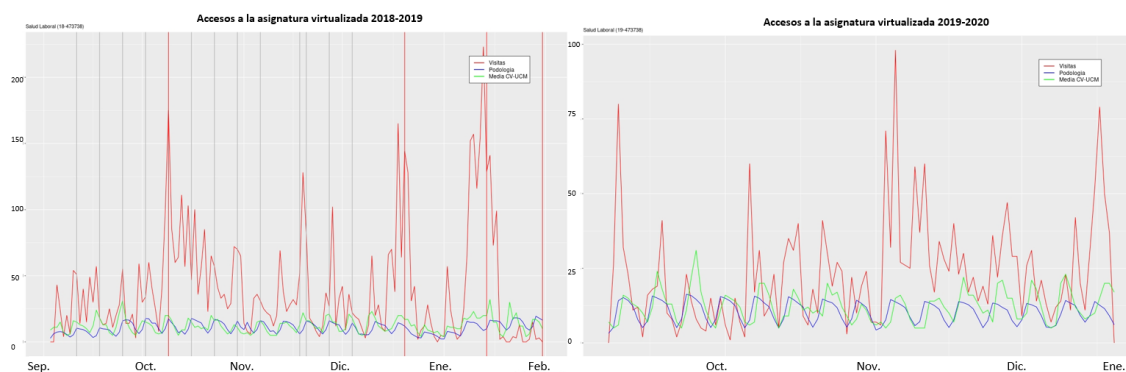


Figura 5. Visitas en los cursos 2018/19 y 2020. En rojo los accesos a SL, en verde la media de accesos al CV de la universidad y en azul los accesos al CV la titulación Grado en Podología.

Si bien la actividad se redujo a la mitad en todas las variables analizadas (visitas, recursos, foros, URLs), las calificaciones se mantuvieron similares. Resulta llamativo el cambio en el comportamiento del grupo de alumnos suspensos en primera convocatoria. Mientras que en 2018/19 hicieron un uso claramente menor del CV que el resto de categorías, en 2019/20 resultó ser el grupo con mayor actividad en cuanto a visitas y accesos a recursos. Si nos centramos en la calificación del examen teórico en vez de en la calificación final (Fig.6), parece que el mayor uso del CV se relaciona con una mejor puntuación en el test, a excepción del grupo de estudiantes con las mejores notas, lo cual podría deberse a que sean personas que necesiten menos apoyo del CV o a que el grupo es poco representativo ($n=2$).

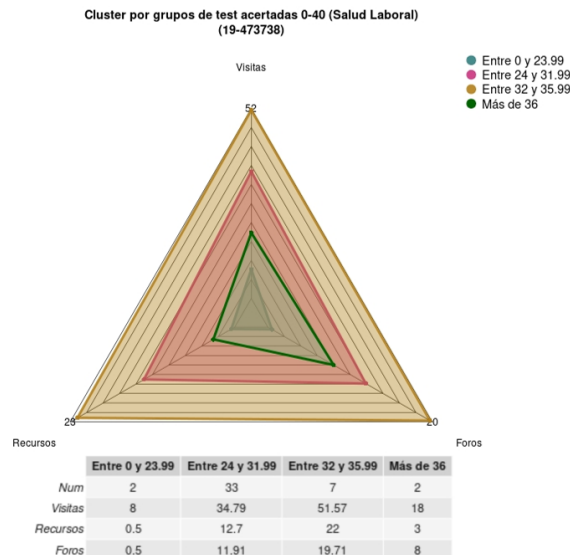


Figura 6. Uso del CV por cada categoría de alumnos de SL en el curso 2019/20 según el número de aciertos en el test.

Por último, se analizó la asignatura E en dos cursos. Se trata de una asignatura de máster cuyo perfil predominante de alumnado es el de mujer enfermera. Es importante señalar que el número de estudiantes pasó de 13 a 31. A pesar del aumento de estudiantes, la actividad en el CV fue mayor en el 2018/19. El tipo de usos registrados en el conjunto de los dos cursos se muestra en la Fig. 7.

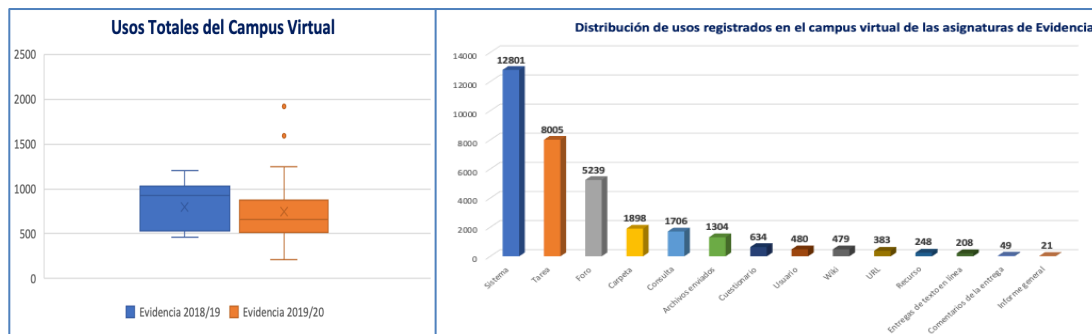


Figura 7. Comparativa de uso del CV en E en 2018/19 vs 2019/20 (dcha.) y distribución de usos totales registrados en el CV en sumando ambos cursos (izqda.)

En cuanto al rendimiento académico en E, no hay grandes variaciones de un curso a otro, lo cual se explica por la metodología empleada. Si bien es interesante señalar que en el examen final hubo mejores resultados en el grupo pequeño, justamente el que tuvo una mayor actividad en el CV y un comportamiento menos disperso. Al relacionar la dispersión entre los usos del CV y la nota final se evidencia una relación lineal positiva baja ($r=0.388$) en el curso 2019/20 mientras que en el 2018/19 el bajo número de alumnos dificulta poder establecer si existe relación (Fig. 8).

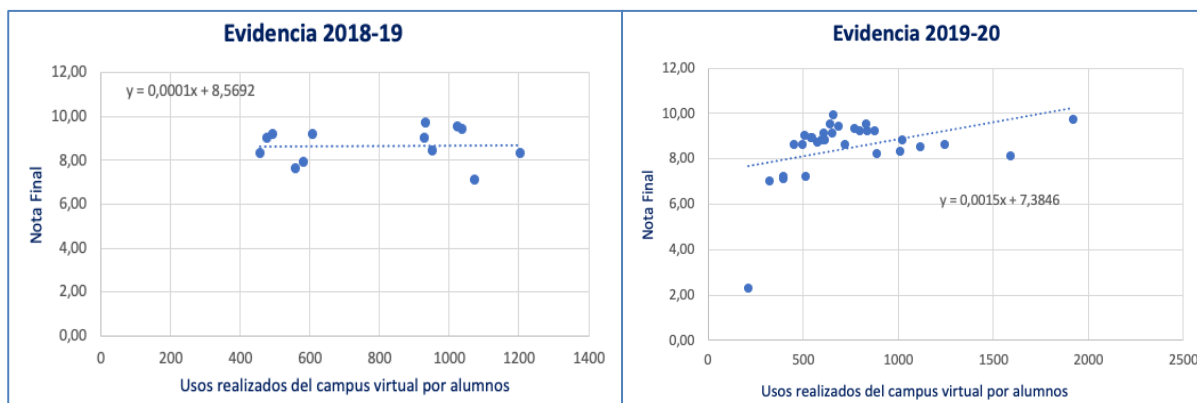


Figura 8. Comparativa de correlaciones entre uso del CV y nota final en E 2018/19 (dcha.) y 2019/20 (izqda.)

3. Análisis de la encuesta realizada a alumnos. La nueva forma de recolectar la encuesta ha supuesto una disminución importante en el número de estudiantes que participan en la misma.
4. Análisis multivariante de los espacios analizados en el PIMCD2018-20 tomados de forma conjunta, por C. Bravo. Nuestro propósito era conocer si, con independencia del espacio, curso, titulación y las características del CV de cada caso, se puede determinar la existencia de clústeres de estudiantes según el uso del CV. El estudio se llevó a cabo sobre 8 espacios y un total del 517 estudiantes. Revela la existencia de 4 clústeres (Fig. 9), de proporciones bastante similares, en función del mayor o menor uso del CV y sus herramientas (accesos a recursos, foros y URLs), y que esos clústeres muestran distintas notas medias en comparación con el global. De tal forma que los individuos de clústeres con mayor uso del CV alcanzan mejores notas en teoría y prácticas de primera convocatoria y nota final en ambas convocatorias que el global (+ 0,5 puntos de media) en tanto que el clúster formado por los estudiantes más pasivos obtuvo peores notas en teoría, prácticas y nota final de primera convocatoria y teoría en segunda convocatoria, siendo la diferencia media respecto del global de -1 punto.

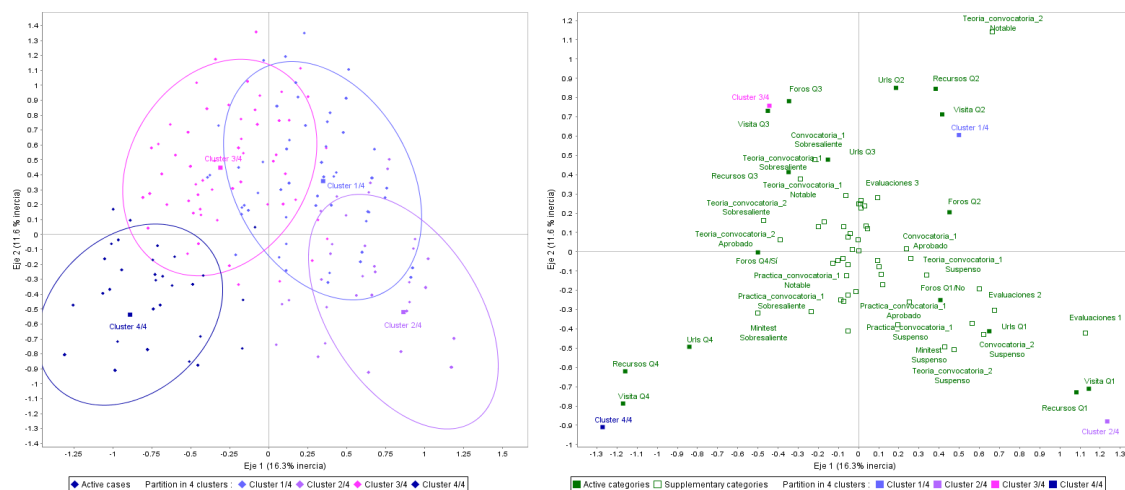


Figura 9. Proyección en los dos primeros ejes factoriales de los individuos y de las variables cualitativas y cuantitativas analizadas

5. Análisis estadístico descriptivo de las asignaturas, correlaciones y árboles de decisión, por C. Bravo. Presentamos tres ejemplos del curso 2019/20. El primero, un árbol que muestra que los estudiantes de BM que hicieron más de 3 autoevaluaciones obtuvieron de media una calificación final superior en 1 punto a los que realizaron 3 o menos (Fig.10).

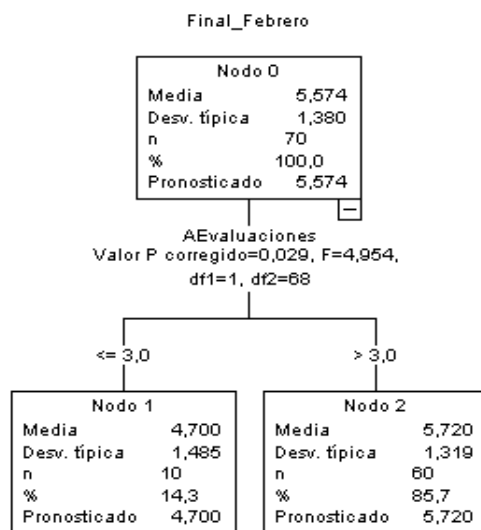


Figura 10. Árbol de decisión realizado según método de crecimiento CHAID (calidad estimada 1,749) para la nota final de BM en primera convocatoria.

A continuación, dos árboles correspondientes a SL (Fig. 11). Se observa que la nota final del estudiante (izqda.) aumenta hasta 2 puntos en los que participan más en foros y hasta 0,5 puntos con el mayor uso de las URLs, mientras que la mayor participación en foros y número de visitas se traduce en 5 y 3 aciertos más, respectivamente, en el examen tipo test (dcha.).

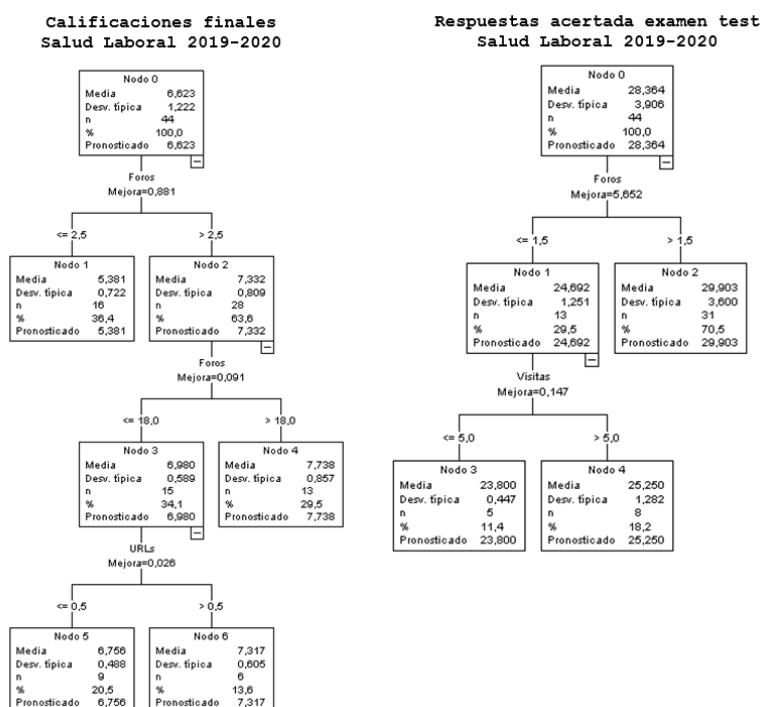


Figura 11. Árboles de decisión realizados para la nota final de SL (izqda.) y nº de aciertos en el examen tipo test en primera convocatoria (dcha.). Método de crecimiento CRT, calidad estimada 0,463 y 9,114, respectivamente.

ANEXO

MANUAL TÉCNICO

Este documento sirve para mostrar la forma de proceder por cualquier usuario, para realizar la evaluación de los logs proporcionados por el Campus Virtual de la UCM (moodle), dentro de la herramienta desarrollada. Esta herramienta hace uso de un script escrito en el lenguaje de programación R, capaz de correr en cualquier tipo de ordenador (Windows, Linux, Mac), que permite recuperar información de distintos documentos y generar gráficos en archivos individualizados en formato jpg así como realizar tratamientos estadísticos complejos. Los distintos archivos que procesar y demás datos de entrada se han “empaquetado” en un documento en formato xml para facilitar la información necesaria para ser procesada. Este archivo puede ser modificado con cualquier editor, aunque sería recomendable utilizar un programa para generarlo de forma automatizada. Este archivo está caracterizado siguiendo una estructura denotada en un archivo DTD (Document Type Definition¹):

```
<!DOCTYPE cursos [  
  <!ELEMENT cursos (curso+)>  
  <!ELEMENT  
    (id,anoaca,author?,title?,degree?,college?,url?,logfile,logvers,studentsfile,gradesfile,fec-ini,fec-  
    fin,grades,parciales?,pruebas?,autoevaluaciones?,foros?,tareas?,groups?)>  
    <!ELEMENT id (#PCDATA)>  
    <!ELEMENT anoaca (#PCDATA)>  
    <!ELEMENT author (#PCDATA)>  
    <!ELEMENT title (#PCDATA)>  
    <!ELEMENT degree (#PCDATA)>  
    <!ELEMENT college (#PCDATA)>  
    <!ELEMENT url (#PCDATA)>  
    <!ELEMENT logfile (#PCDATA)>  
    <!ELEMENT logvers (#PCDATA)>  
    <!ELEMENT studentsfile (#PCDATA)>  
    <!ELEMENT gradesfile (#PCDATA)>  
    <!ELEMENT fecini (#PCDATA)>  
    <!ELEMENT fecfin (#PCDATA)>  
    <!ELEMENT grades (main_grade,main_grade2?,grade+)>  
    <!ELEMENT main_grade (name,passgrade?)>  
    <!ELEMENT main_grade2 (name,passgrade?)>  
    <!ELEMENT grade (name,passgrade?)>  
    <!ELEMENT parciales (parcial*)>  
    <!ELEMENT pruebas (prueba*)>  
    <!ELEMENT autoevaluaciones (autoevaluacion*)>  
    <!ELEMENT foros (foro*)>  
    <!ELEMENT tareas (tarea*)>  
    <!ELEMENT groups (group*)>
```

¹

Document Type Definition <https://www.w3.org/QA/2002/04/valid-dtd-list.html>

```

<!ELEMENT group (ingrade,cod_split,min_split,max_split)>
<!ELEMENT ingrade (#PCDATA)>
<!ELEMENT cod_split (#PCDATA)>
<!ELEMENT min_split (#PCDATA)>
<!ELEMENT max_split (#PCDATA)>
<!ELEMENT parcial (#PCDATA)>
<!ELEMENT prueba (#PCDATA)>
<!ELEMENT autoevaluacion (#PCDATA)>
<!ELEMENT foro (#PCDATA)>
<!ELEMENT tarea (#PCDATA)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT passgrade (#PCDATA)>
]>

```

El archivo XML (*eXtensible Markup Language*²) actúa como índice y contiene la información que precisa el aplicativo para gestionar los datos y elaborar los resultados y estadísticas:

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE address SYSTEM "index.dtd">
<cursos>
  <curso>
    <id>19-107910</id>
    <anoaca>2019/20</anoaca>
    <author>Yolanda Fuentes Peñaranda</author>
    <title>Evidencias MUICS</title>
    <degree></degree>
    <college>126</college>
    <url></url>
    <logfile>./19-107910 Registros.csv</logfile>
    <logvers>2.9.1</logvers>
    <studentsfile>./19-107910 Alumnos.csv</studentsfile>
    <gradesfile>./19-107910 Calificaciones.csv</gradesfile>
    <resourcesfile></resourcesfile>
    <fecini>2019-09-09</fecini>
    <fecfin>2019-12-20</fecfin>
    <grades>
      <main_grade>
        <name>Nota_Final</name>
        <passgrade>5</passgrade>
      </main_grade>
      <grade>
        <name>Total_Actividades</name>
        <passgrade>2</passgrade>

```

```

        </grade>
    </grades>
    <parciales>
        <parcial>2020-01-09</parcial>
    </parciales>
</pruebas></prueba>
    <autoevaluaciones></autoevaluaciones>
    <foros></foros>
    <tareas>
        <tareas>2019-09-30</tareas>
        <tareas>2019-11-25</tareas>
        <tareas>2019-12-19</tareas>
        <tareas>2020-01-27</tareas>
    </tareas>
    <groups>
        <group>
            <ingrade>Nota_Final</ingrade>
            <cod_split>Entre 0 y 4,99</cod_split>
            <min_split>0</min_split>
            <max_split>5</max_split>
        </group>
        <group>
            <ingrade>Nota_Final</ingrade>
            <cod_split>Entre 5 y 6,94</cod_split>
            <min_split>5</min_split>
            <max_split>6.95</max_split>
        </group>
        <group>
            <ingrade>Nota_Final</ingrade>
            <cod_split>Entre 6,95 y 8,54</cod_split>
            <min_split>6.95</min_split>
            <max_split>8.55</max_split>
        </group>
        <group>
            <ingrade>Nota_Final</ingrade>
            <cod_split>Más de 8,54</cod_split>
            <min_split>8.55</min_split>
            <max_split>1000</max_split>
        </group>
    </groups>
</curso>
</cursos>

```

1. El curso o cursos que evaluar:<curso>

- a. <id>: Identificación dentro del Campus Virtual
- b. <anoaca>: Año académico (20XX/XX). Se mostrará en la cabecera de los gráficos
- c. <author>: Nombre del Profesor/a de la asignatura
- d. <title>: Descripción del Curso
- e. <degree>: Grado o titulación al que pertenece la asignatura
- f. <college>: Centro al que pertenece el profesor/a.
- g. <url>: URL del curso en el Campus Virtual
- h. <logfile>: Fichero con el log de actividad del moodle. Extraído desde la propia herramienta moodle
- i. <logvers>: Versión de logs proporcionados por moodle. (2.6 /2.9)
- j. <studentsfile>: Fichero con los datos de los alumnos desde GEA.
- k. <gradesfile>: Fichero de calificaciones extraído desde moodle.
- l. <resourcesfile>: Fichero con el mapa de los recursos disponibles dentro del curso.
- m. <fec-ini>: Fecha de inicio del curso
- n. <fec-fin>: Fecha de finalización del curso
- o. <grades>: Descripción de las calificaciones empleadas en la asignatura. El fichero de calificaciones <logfile> puede contener más calificaciones, pero solo las indicadas en este punto serán consideradas. Cada asignatura dispondrá de dos calificaciones finales (<maingrade> y <maingrade2> y hasta 30 calificaciones parciales que contemplen distintas actividades (parciales, autoevaluaciones, participación en foros, wikis, blogs). Para cada una de estas calificaciones debe detallarse un nombre (que deberá coincidir con el indicado en el fichero <gradesfile>) y una escala de valoración.
- p. <parciales>: Fechas de realización de los exámenes parciales
- q. <pruebas>: Fechas de realización de distintas pruebas evaluativas externas a moodle
- r. <autoevaluaciones>: Fechas de realización de ejercicios de autoevaluación (evaluable).
- s. <foros>: Fechas de entrega de las participaciones en foros (evaluable).
- t. <tareas>: Fechas de entrega de las tareas (evaluable).
- u. <groups>: División de calificaciones para cada nota que quiera ser evaluada. Permite realizar comparativas de uso del Campus Virtual en función de las calificaciones obtenidas en cada nota según los tramos indicados.

2. El archivo <logfile> contiene toda la información registrada en los ordenadores con la actividad de los alumnos y profesores. Habitualmente esta información es utilizada para detectar errores o mal funcionamiento de los módulos que componen el campus virtual, pero en este caso nos interesa para trazar la actividad de los usuarios dentro de la asignatura. En estos archivos no solo figura la información de accesos a las herramientas, también la actividad que se realiza sobre las propias herramientas. Por ejemplo, nos puede indicar el número de veces que un alumno realiza una tarea, pero también si ha sido corregida, visualizada por otros alumnos y/o profesores, el número de archivos que ha adjuntado en esa tarea, Además, en estos archivos figura la información de accesos de cualquier usuario (Profesor y Estudiante) dentro del curso. Por este motivo, es necesario realizar un filtro para procesar la información contenida en estos archivos. El contenido de estos archivos es:

id	time	userid	ip	course	module	cmid	action	url	info	idUs	DNI	nombre	apellidos	idEspacio	rol
467237550	1440412845	68801	17683114157	58232	course		0 view	view.php?id=58232	58232	5196	50306370	MARIA DEL PILAR	ALVAREZ VAZQUEZ	15-93062	Profesor
467238606	1440413691	68801	17683114157	58232	course		0 view	view.php?id=58232	58232	5196	50306370	MARIA DEL PILAR	ALVAREZ VAZQUEZ	15-93062	Profesor
467258644	1440436696	68801	95124095	58232	course		0 view	view.php?id=58232	58232	5196	50306370	MARIA DEL PILAR	ALVAREZ VAZQUEZ	15-93062	Profesor
467258669	1440436712	68801	95124095	58232	resource	1811862	view	view.php?id=1811862	1282173	5196	50306370	MARIA DEL PILAR	ALVAREZ VAZQUEZ	15-93062	Profesor

La estructura cambia con la versión de moodle, pero, básicamente, nos fijamos en el módulo utilizado, la acción, y el rol del usuario para determinar la actividad de los usuarios dentro del curso. También es posible obtener estos datos desde la propia herramienta de Moodle como informe y extraerla en formato CSV (Comma-Separated Value³).

El archivo permite evaluar varias asignaturas simultáneamente. El número de Parciales, Autoevaluaciones, Foros y Tareas es ilimitado.

Las fechas de inicio y fin no tiene por qué coincidir con las fechas de inicio y fin del curso. Pueden concretarse a un periodo para estudiar en profundidad ese periodo de tiempo.

El campo loggers puede contener tres valores distintos: 2.6, 2.9 puesto que ambas versiones tienen estructuras distintas. También puede ser 2.9.1 si el fichero log se extrae desde la herramienta de Moodle mediante informes.

3. El archivo <studentsfile> debe contener los siguientes campos:

- a. N° de Matrícula
- b. N° de Convocatoria
- c. Procedencia (Opcional)
- d. Sexo
- e. Apellidos
- f. Nombre
- g. DNI/NIF/Pasaporte

4. El archivo <gradesfile> contiene las calificaciones obtenidas en cada uno de los parciales, así como las notas de las convocatorias de junio y julio (o febrero). Cada nota debe ir identificada con un título que deberá coincidir con las notas marcadas dentro de la estructura de <grades>

5. El archivo <resourcefile> indica información sobre los recursos, su identidad dentro de Moodle, el formato del recurso (HTML, imagen, PDF, vídeo, ,,,)

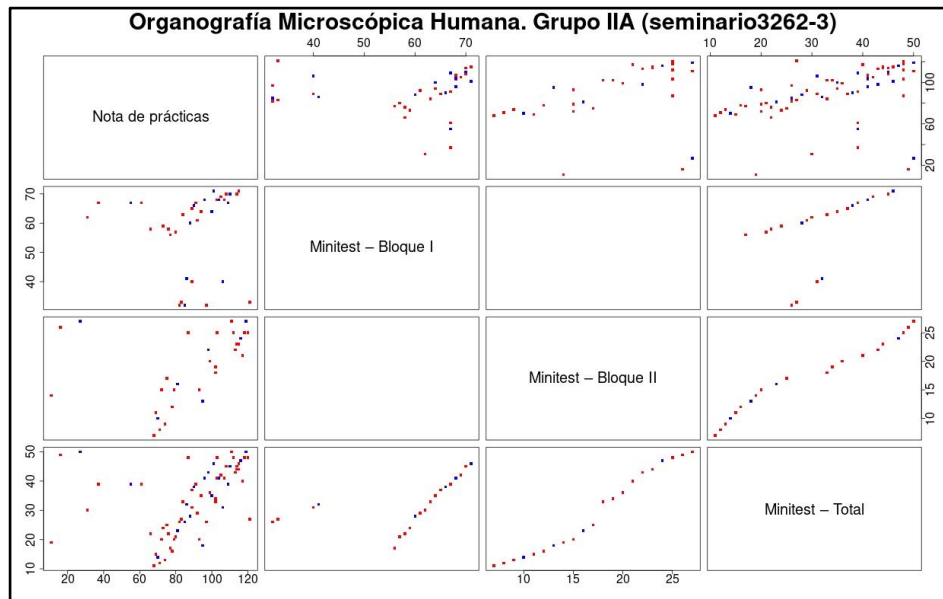
Una vez preparados los archivos, la ejecución de la aplicación obtendrá distintos gráficos en formato jpg en archivos separados. Los gráficos que se obtienen son los siguientes:

1. Utilización general de las herramientas en cada curso. Describe, de forma global, que tanto por ciento de estudiantes acceden al curso y a cada una de las actividades.

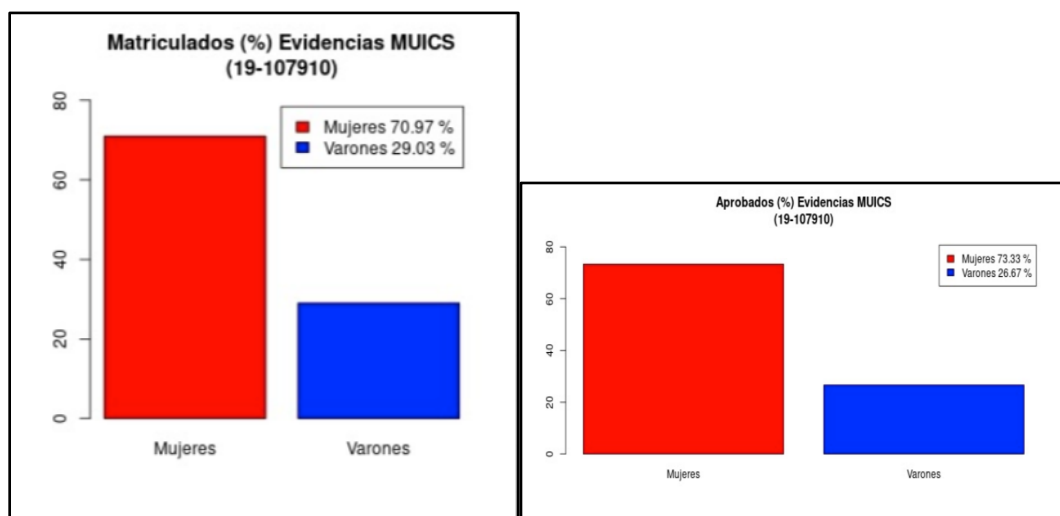
id	title	Visitas	Recursos	AEvaluaciones	Foros	Tareas	URLs	Glosarios	Wikis
17-190652	Organografía Microscópica Humana. Grupo IA	98.75	98.75	0	77.50	11.25	62.50	0	0
seminario3262-3	Organografía Microscópica Humana. Grupo IIA	83.12	75.32	0	37.66	0.00	0.00	0	0
seminario-invest-5914-19	seminario organo IA	100.00	100.00	0	52.63	12.63	44.21	0	0

³ Comma-Separated Value <https://tools.ietf.org/html/rfc4180>

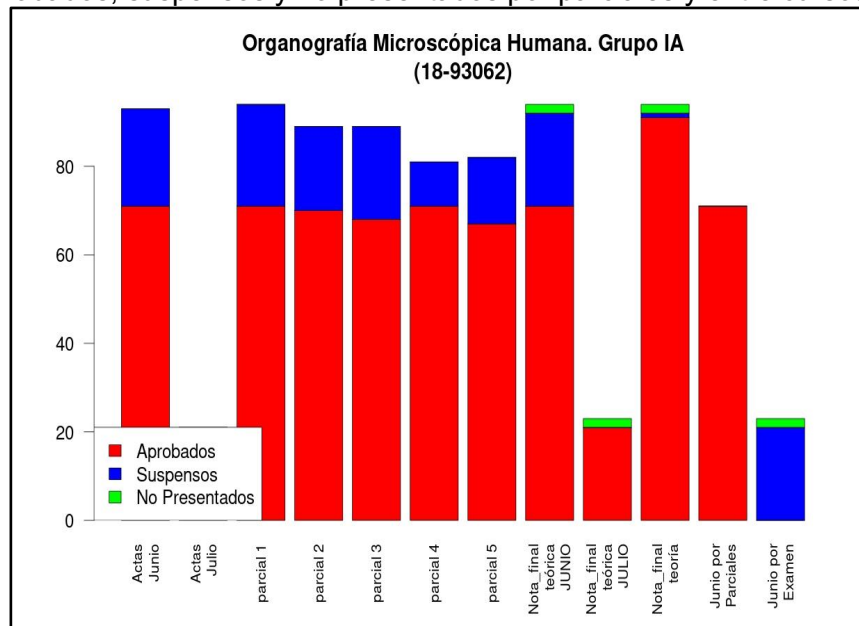
- Distribución de entre pares de calificaciones. Este gráfico tridimensional muestra la relación de los estudiantes con respecto a las calificaciones entre parciales. Muestra en distinto color los alumnos por sexo para valorar cualquier relación. Es útil para conocer si un grupo de alumnos ha podido perder la progresión durante el curso.



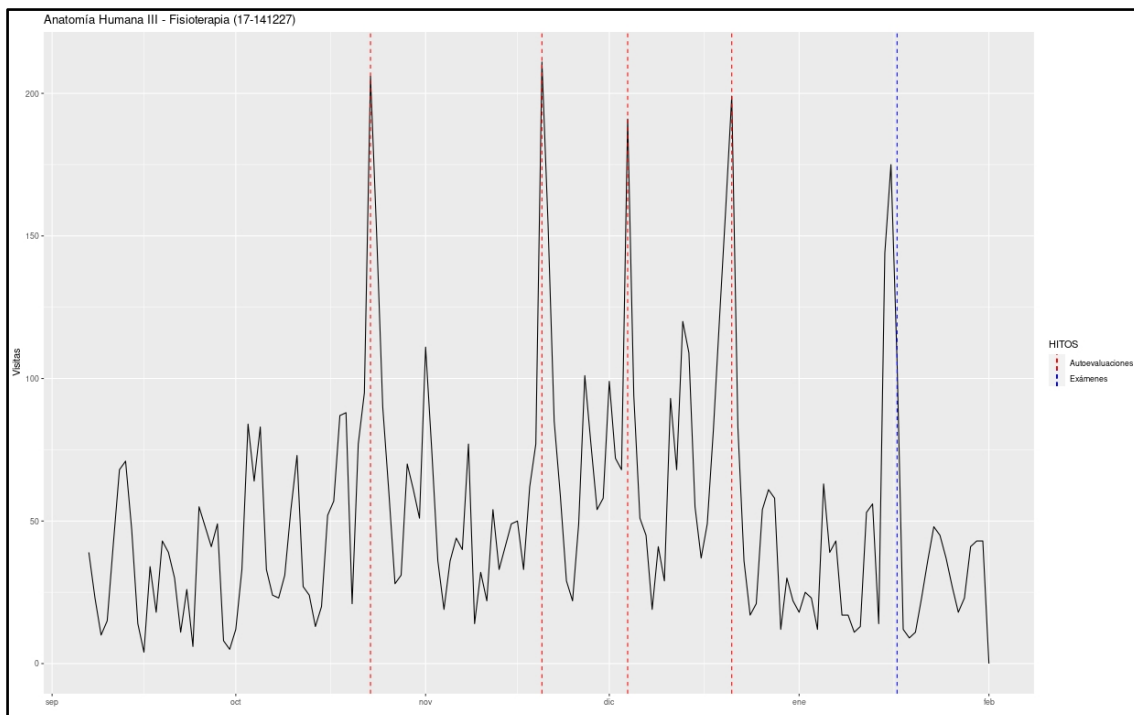
- Estudiantes matriculados por sexo. Permite conocer la variación del número de estudiantes matriculados por sexo entre distintos cursos y el tanto por ciento de aprobados por sexo.



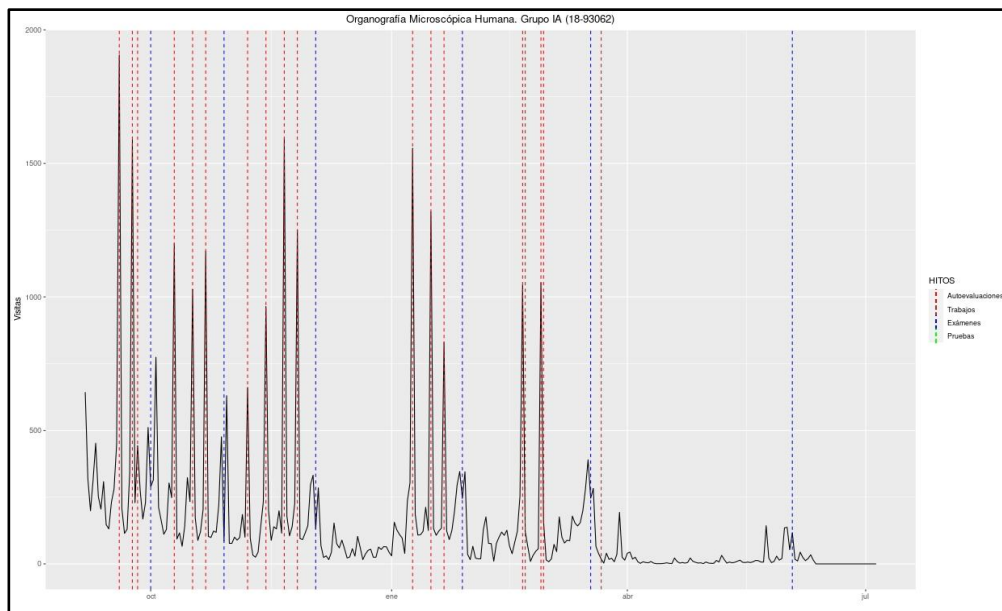
4. Distribución de calificaciones por parciales. Permite comparar la evolución del número de aprobados, suspensos y no presentados por parciales y entre cursos.



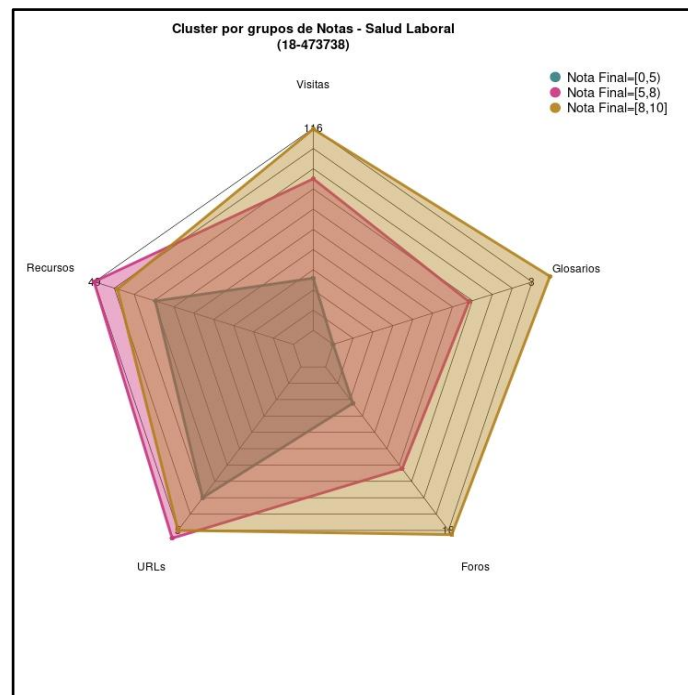
5. Accesos al curso por fechas, Para cada curso indicado, se muestra una gráfica de accesos por fecha con referencia a las fechas de los parciales, pruebas externas, autoevaluaciones, fechas de entrega de tareas o participación en foros si hubieran sido indicadas.



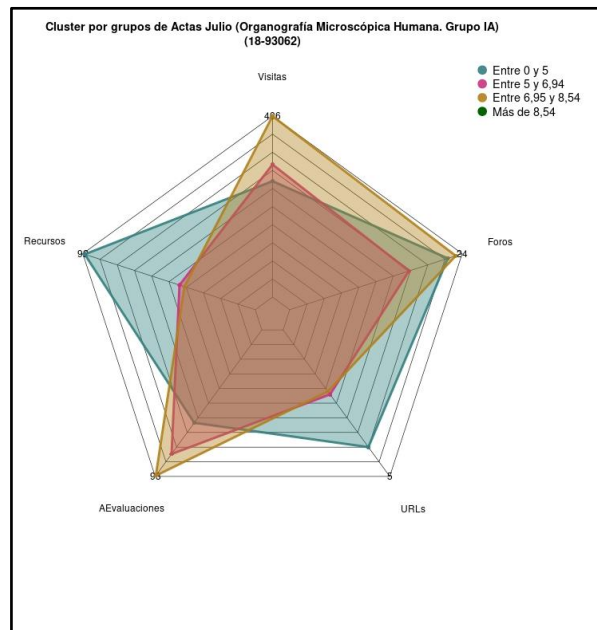
6. Acceso a actividades dentro del curso. Para cada una de las actividades propuestas (Wikis, Autoevaluaciones, Glosarios, Accesos URL externas, Foros, Tareas y Recursos). Muestra, por fechas, las actividades realizadas dentro del curso. Al igual que la anterior gráfica, muestra como referencia las fechas de los parciales y las autoevaluaciones y la media de recursos descargados por día durante el periodo estipulado (<fec-ini> y <fec-fin>).



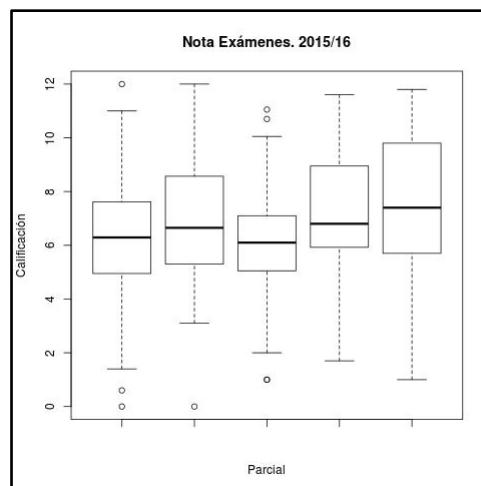
7. Clúster de estudiantes por nota obtenida. Segmenta los resultados finales en tres rangos (de 0 a 5, de 5 a 8 y de 8 a 10), mostrando, mediante un diagrama de radar o estrella, el uso de las herramientas que realiza cada clúster de estudiantes.



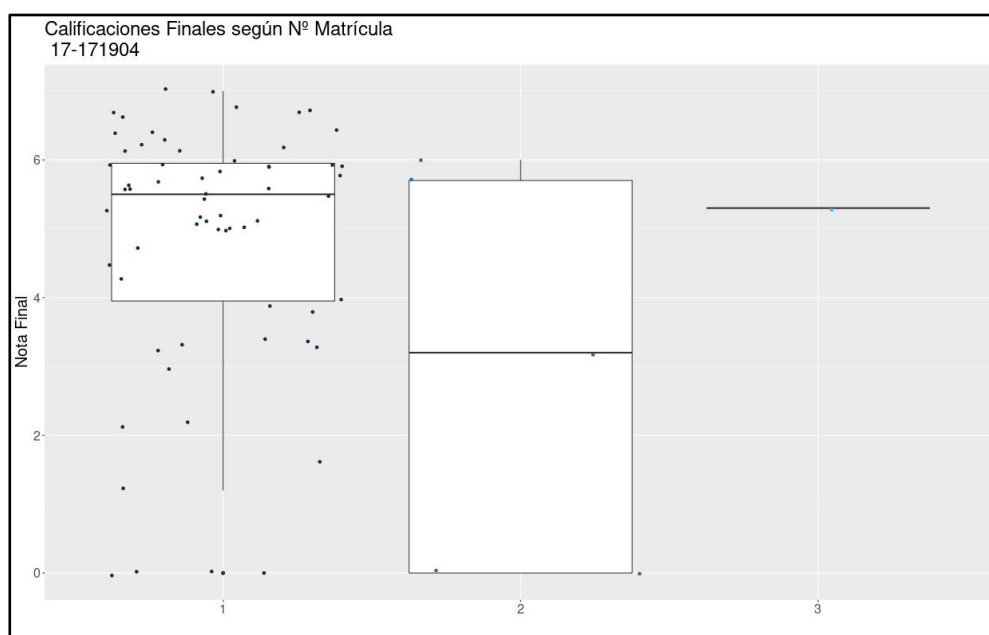
8. Diagrama de red para cada calificación según los tramos especificados. Para cada calificación a valorar se pueden establecer tramos de notas y evaluar el comportamiento de los alumnos en las distintas actividades según la nota obtenida.



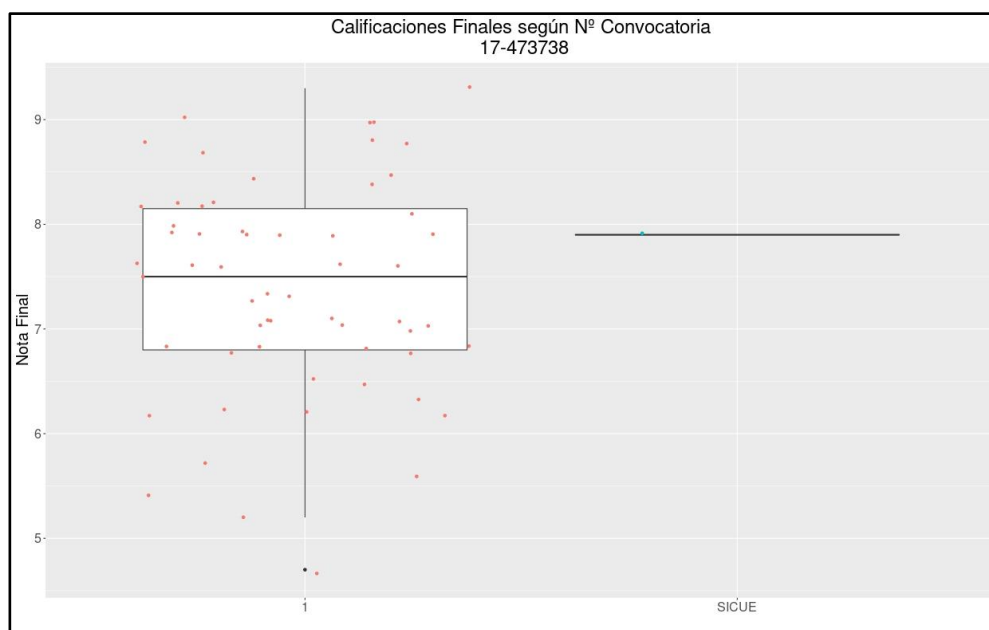
9. Diagrama de cajas por parciales. Permite observar en una sola gráfica la distribución de las calificaciones por cada parcial (cuartiles, medias, rango intercuartílico y valores atípicos). Con este gráfico podemos identificar alumnos con calificaciones excepcionales en ambos sentidos y comparar los valores entre los distintos parciales.



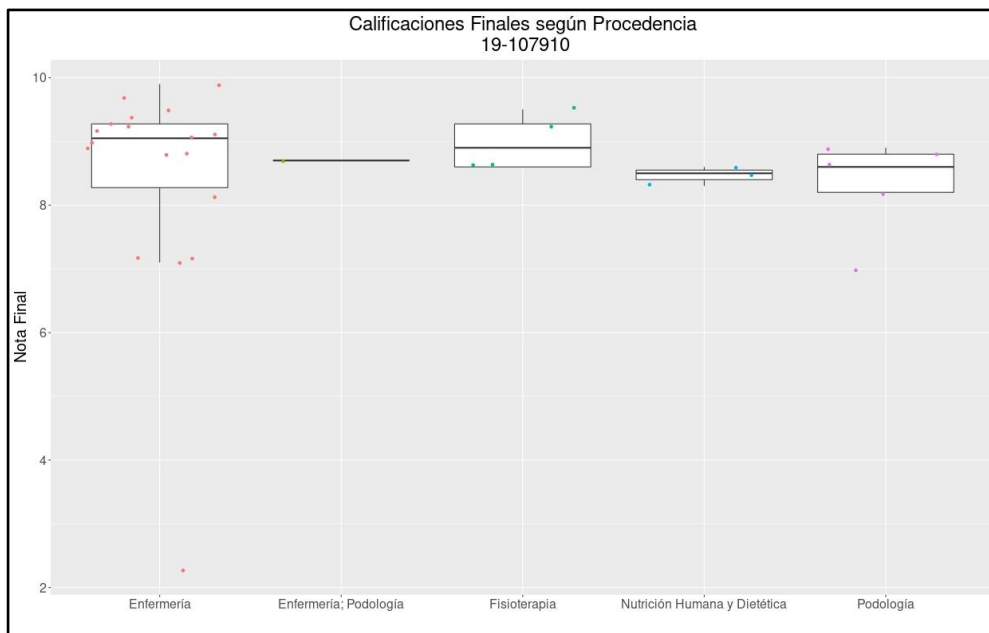
10. Calificaciones obtenidas por número de matrícula.



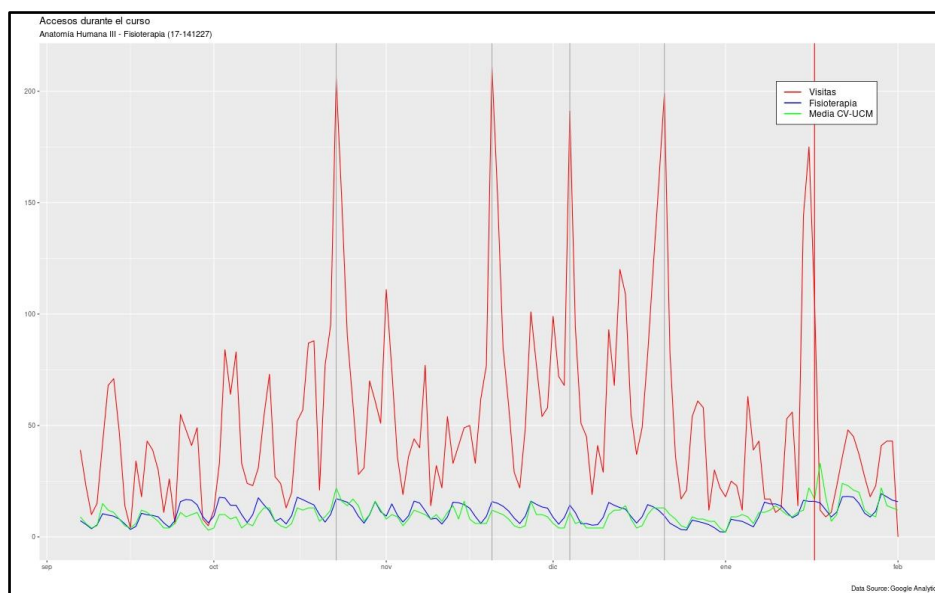
11. Calificaciones por número de convocatoria



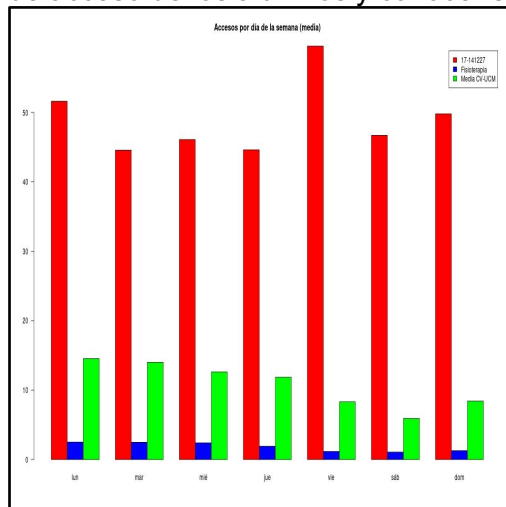
12. Calificaciones por procedencia del estudiante



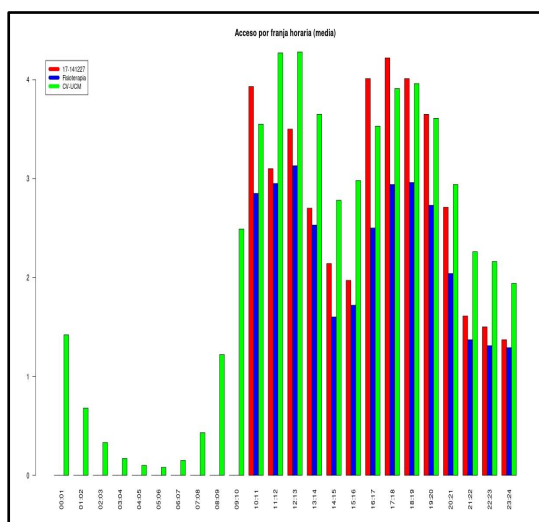
13. Comparativa de accesos con la media de las asignaturas virtualizadas en su titulación y con la media de asignaturas del CV de la UCM.



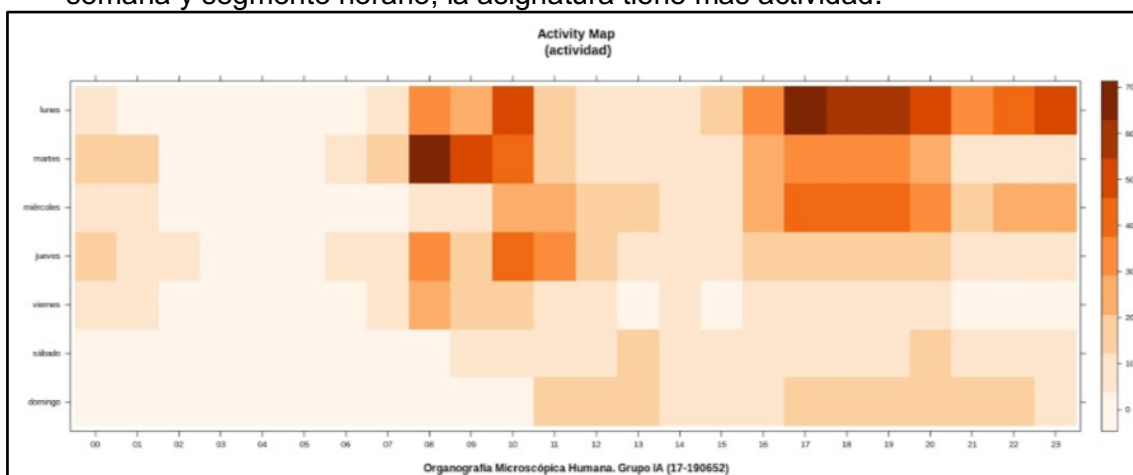
14. Accesos por día de la semana. Este gráfico permite valorar los accesos de los usuarios por día de la semana y compararlo con los accesos del resto de asignaturas dentro del Campus Virtual de la universidad. Este gráfico, junto con el siguiente, permite identificar patrones de acceso de los alumnos y conocer sus ritmos de trabajo.



15. Accesos a la asignatura por franja horaria. Permite conocer a qué hora se conectan con más frecuencia los estudiantes al Campus Virtual y programar actividades acordes a esos ritmos.



16. Mapa de Calor de la asignatura. Muestra un gráfico para conocer en qué días de la semana y segmento horario, la asignatura tiene más actividad.



17. Árboles de decisión sobre la nota final. Los árboles de decisión nos ayudan a tomar la decisión “más acertada”, desde un punto de vista probabilístico, ante un abanico de posibles decisiones. En este caso sirve para evaluar cómo afecta probabilísticamente la participación de cada alumno en cada una de las herramientas disponibles en el curso de cara a la nota final obtenida.

